HEALTH ECONOMICS
Health Econ. **8**: 459–472 (1999)

HEALTH ECONOMICS AND ECONOMETRICS

SENSITIVITY OF ELASTICITY ESTIMATES FOR OECD HEALTH CARE SPENDING: ANALYSIS OF A DYNAMIC HETEROGENEOUS DATA FIELD

JENNIFER ROBERTS*
Sheffield Health Economics Group, ScHARR, University of Sheffield, Sheffield, UK

SUMMARY

Current econometric research on the determinants of health care spending is characterized by three serious shortcomings: a lack of attention to dynamics, a failure to deal with heterogeneity and a neglect of sensitivity testing. The latter shortcoming is a particular problem in an area where there is little theoretical guidance with which to judge the validity of estimates. This paper deals with these three problems by adopting recently developed techniques for analysing dynamic heterogeneous data fields containing non-stationary variables. This method allows for parameter heterogeneity across countries and is preferred to pooling with country specific effects, since the latter does not yield consistent estimates in this context, despite its common use in studies of health care spending. Long-run and short-run elasticity estimates are found to be sensitive to both model specification and sample composition. An attempt is made to compensate for the weak theoretical base by increasing the emphasis on statistical and econometric criteria for model evaluation, and by subjecting the results to rigorous sensitivity testing. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS — health care expenditure; data fields; dynamic heterogeneous panels

INTRODUCTION

Joseph Newhouse's claim, in 1977, that health spending was a luxury good, has prompted over 20 years of empirical research into the determinants of health care expenditure [1]. During this time there has been little progress beyond the finding that variations in per capita national income are closely correlated with variations in per capita health spending. In particular, there is still considerable debate regarding the size of the income elasticity of health spending [2]. The failure to progress knowledge fuels criticism that this type of macro analysis is not helpful in answering the important questions about health care expenditure, and in particular in contributing to policy advice.

The proliferation of econometric work on health care spending is characterized by three serious shortcomings. Firstly, a lack of attention to the dynamics of the relationship between health spending, income and other potential regressors. Secondly, a failure to adequately deal with the heterogeneity of relationships across countries. Thirdly, a neglect of diagnostic and sensitivity testing, and a failure to systematically consider the robustness of the results. The latter shortcoming is a particular problem in an area where there is little theoretical guidance against which to judge estimates.

* Correspondence to: Sheffield Health Economics Group, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. E-mail: J.R.Roberts@sheffield.ac.uk

CCC 1057–9230/99/080459–14$17.50
Copyright © 1999 John Wiley & Sons, Ltd.
Received 29 September 1998
Accepted 12 January 1999

This paper contributes to knowledge in this area by analysing the determinants of health care spending as a dynamic, heterogeneous data field. 'Data field' is a term coined by Quah [3] to describe a panel data set consisting of a relatively small number of cross sections, each with a relatively long time series dimension. Like the health expenditure data, the observations are often measured at country level. This can be contrasted with traditional panels measured at the level of the individual, which have large numbers of cross-sectional observations measured at very few points in time. It should also be noted that unlike many traditional panels (for example the British Household Panel Survey), data fields are not explicitly designed as a panel to analyse a specific set of issues. Each country collects its data independently of the others and there are questions of comparability [4].

Data fields provide extremely sophisticated data sets. The time series element enables examination of dynamic relationships and the combination of cross-section variation and long time series allow us to consider parameter heterogeneity over the cross-section groups. As well as being of value to anyone interested in the macroeconomics of health care spending, this paper also has a more general role in illustrating recent developments in the econometric analysis of data fields. Like many variables that form the time series in a data field, the variables here are non-stationary. These features warrant particular consideration, and up to now there has been a shortage of empirical applications to illustrate the theoretical developments.

The first two sections of this paper consider why the relationships between health spending, income and a number of other potential regressors are dynamic and heterogeneous. The third section considers the appropriate estimation techniques for dynamic, heterogeneous relationships involving non-stationary variables. Here we see that pooling the data, even after allowing for fixed effects, does not result in consistent parameter estimates in the dynamic case.

The data, method and initial results are outlined in the fourth section. In the fifth section these results are subject to sensitivity testing. This reveals that long-run elasticity estimates are sensitive to model specification, particularly to the inclusion of a time trend. It is difficult to obtain robust estimates of short-run effects, but these are important and appear to be consistent with a long-run equilibrium relationship between health spending and income. A concluding section summarizes the findings and discusses the implications for policy and future research.

HETEROGENEITY IN HEALTH CARE SPENDING

Early cross-section studies of health care spending have attracted a number of criticisms of which Parkin *et al.* [5] and Gerdtham [6] provide a good discussion. Of current relevance are those criticisms relating to the imposition of unrealistic assumptions regarding the homogeneity of tastes, preferences and production functions across countries. Health systems differ across countries according to economic, social, political and cultural factors; as well as beliefs about medicine and health. In an econometric framework, the problem centres on the non-observability of many of these effects, which are likely to be correlated with the observed regressors, and therefore lead to specification bias in the estimated coefficients.

Methods of dealing with parameter heterogeneity are still underdeveloped in the health care expenditure literature. Consider a general model of the form:

$$h_{it} = g(\beta' \mathbf{x}_{it}) + \varepsilon_{it} \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T. \tag{1}$$

where h_{it} is total health expenditure per capita for country i in time t, and $\mathbf{x}$ is a vector of regressors. This vector always includes income per capita, but the other columns vary across the studies in question, including such things as the proportion of the population aged over 65, the proportion of public financing of health care and the relative price of health care services. While the models are essentially *ad hoc*, the choice of right hand side variables is influenced by the vast discursive literature on the possible determinants of health care spending [7–9]. The log–linear functional form, g, is popular because the coefficient estimates can be interpreted as elasticities. Income and expenditure data for the different countries are generally adjusted using purchasing power parities.

If t is a single point in time, model (1) represents a cross section, and implicitly imposes the homogeneity assumptions. This is unrealistic and imposes invalid restrictions on the parameters of

Copyright © 1999 John Wiley & Sons, Ltd.

model (1). Heterogeneity is present across countries, despite the fact that almost all of the empirical work is limited to the OECD. Early attempts to account for heterogeneity involved the inclusion of dummy variables to reflect various institutional characteristics. For example, Leu [10] included dummy variables for the centralized national health systems of the United Kingdom and New Zealand, and estimated that health expenditure in these systems was 20% lower than the other OECD countries, even after income and demographic composition were taken into account.

This method of allowing for varying intercepts extends to the panel data framework that has recently been adopted in the analysis of health care spending. Rather than specifying dummy variables a priori in what can be a rather *ad hoc* fashion, the method allows for country specific intercepts. The errors from model (1) are defined as:

$$\varepsilon_{it} = \mu_i + u_{it}, \tag{2}$$

where μ_i is a country specific fixed effect term, and u_{it} can be thought of as the usual stochastic error term [11].

Some pooled studies fail to test Equation (1) against this more general fixed effect model [12,13]. This casts doubt on their results, since a number of studies have shown that the pool restriction is invalid, and that inclusion of fixed effects makes a significant difference to the estimated slope coefficients [6,14–16].

Only a very small number of studies have gone beyond country specific intercepts, and attempted to deal with the possibility of slope heterogeneity. Pfaff [17] divides the OECD countries into two groups according to the mode of finance (mainly taxed-based or mainly contributory financed systems), and finds that the income elasticities of health care spending are generally higher in the tax-financed systems. O'Connell [15] uses country specific demographic variables and finds that the effect of the ageing population varies substantially across the OECD countries. Gerdtham [6] takes a systematic look at the effect of sample composition by dropping each country in turn, and finds significant parameter variation, especially in relation to the income elasticity of health care spending.

Analysing individual country time series is an extreme method of accounting for parameter heterogeneity. Both Murillo *et al.* [18] and Saez and Murillo [19] rule out the use of pooled data because of the invalidity of the homogeneity assumptions. However, this type of analysis does not take full advantage of the information provided by data fields such as the OECD health data set. The individual time series solution results in small sample sizes and does not allow us to estimate the *average* affect of say income on health spending across the OECD.

THE DYNAMICS OF HEALTH CARE SPENDING

Both the cross section and early panel studies of health spending assumed that all effects of income, demographics, relative price, etc., occurred in the current period. However, there are a number of reasons to expect lags in the effects of various variables. Consumption theory suggests that health expenditure will relate to both permanent and transitory components of income. The actual process of expenditure adjustment is not well understood and depends on many factors including organizational dynamics, technological change and expectations [20]. Public expenditure on health care is likely to be influenced by electoral cycles. Expenditure, particular on capital goods, is committed over long periods. Labour costs are a high proportion of total health spending and it has been shown that institutional rigidities in the US health system, delay adjustment of health services employment to macroeconomic conditions and other shocks [21]. This tendency is likely to be even greater in other OECD countries with a higher proportion of health employment in the public sector.

In the recent literature consideration of dynamics in health spending has been tied up with the issue of non-stationarity, and the possibility of cointegrating relationships. In summary the regression of non-stationary variables is likely to result in a spurious regression problem, by which the apparent close correlation between health care expenditure and, say, per capita income is the result of common trends rather than evidence of any real economic relationship. Non-stationarity can arise from deterministic and/or stochastic trends in the data, and if variables possess a stochastic trend component they should be differenced to impose stationarity and to avoid

Copyright © 1999 John Wiley & Sons, Ltd.

misleading inference; these variables are said to be integrated of order one, or $I(1)$. In the special case where a linear combination of non-stationary variables is itself stationary, then this combination represents the cointegrating, or long-run, relationship, which can be specified in levels, with short-run dynamics modelled via an error correction process [22].

There is substantial evidence that many of the series commonly used to investigate the determinants of health care spending are non-stationary [14,16,23]. A recent paper by McCoskey and Selden [24] provides contradictory evidence. However, their findings are questioned by Hansen and King [25], and turn out to be very sensitive to the exclusion of a time trend from their testing equations; an issue which is discussed in more detail in the results section below. The evidence on cointegration is less conclusive [14,16,18,19,23]. The contradictory evidence is hardly surprising given the relatively short time series employed, and the many methods of testing for cointegration. It is also questionable, how useful the tests are in the current context. Intuitively there must be a long-run relationship between health spending and income, with a long-run elasticity of at most unity. Studies that focus simply on detecting a cointegrating relationship are only concerned with long-run inference, and neglect short-run dynamics, which are likely to be more relevant to policy formulation. The dynamic structure of the models is simplified, and many of the studies have had to allow for serially correlated errors in order to cope with omitted dynamics [13,15,16].

A small number of papers have considered short-term effects by estimating error correction models (ECMs) [6,18,19]. Gerdtham [6] estimates a first order autoregressive distributed lag (ADL) model with fixed effects, for a 22 country OECD panel, 1970 to 1987. He tests a variety of restrictions of this general model and prefers an ECM with long-run restrictions on the age and finance variables and a zero restriction on price effects. This model imposes a unitary long-run income elasticity on health spending, and his estimates suggest that health spending adjusts towards its equilibrium value at a rate of approximately 22% per year, although this estimate is sensitive to sample composition, illustrating the importance of parameter heterogeneity.

ANALYSIS OF DYNAMIC HETEROGENEOUS DATA FIELDS

The panel econometrics literature tends to focus on intercept variation, with little attention paid to the possibility of variation in slope coefficients across the groups involved. This may be due to the fact that when the main estimators were developed, most available panels only had a relatively short time dimension, hence, it was not possible to investigate slope heterogeneity. However, it has also been suggested that slope heterogeneity does not matter when we are interested in the *average* effect of an exogenous variable on the dependent variable [26], and further that long-run equilibrium responses are less likely to be subject to slope heterogeneity than short run effects. While the former is true in the static case with strictly exogenous regressors and coefficients varying randomly across groups, it does not extend to dynamic models containing a lagged dependent variable. Substantial biases can occur in the estimate of the average long-run effect when a common slope is incorrectly imposed on a heterogeneous model.

Two recent papers discuss the appropriate estimation procedures for dynamic heterogeneous panels, such as the OECD health expenditure data field [27,28]. To summarize the main findings, if both N and T are large there are four alternative procedures for estimating the average effect of some exogenous variable (say per capita GDP) on a dependent variable (per capita health spending). Firstly, the mean group estimator, in which separate time series regressions are estimated for each group and the coefficients are averaged across groups. Secondly, the pooled estimator in which data is pooled by imposing common slopes and allowing for country specific intercepts, as in Equation (2). Thirdly, aggregate time series regressions on the data averaged across countries. Finally, cross-section regressions with the data averaged over time.

In the static case with strictly exogenous regressors and coefficients varying randomly across groups, all four procedures provide unbiased, consistent estimates of the average effects. However, in the dynamic case only the mean group and cross-section estimators are consistent. This is true even if the underlying variables are $I(1)$, as long as the individual group relationships are cointegrated. It should be noted that for the

Copyright © 1999 John Wiley & Sons, Ltd.

cross-section estimator the data must be averaged over relatively long time periods. The early cross-section work on health expenditure, with data from just 1 or 2 years [1,5,10], does not yield consistent estimates.

The aggregate time series estimator has not been used in health expenditure studies and will not be discussed further here. The pooled estimator has been commonly employed in recent work, but pooling fails to produce consistent estimates in the dynamic case, even if country specific effects are allowed for. The results of Gerdtham's fixed effects estimation for the ECM [6] should therefore be treated with caution, since this estimator tends to underestimate the average short-run effect and overestimate the average long-run effect [27]. A further note of caution is that under slope heterogeneity with $I(1)$ data the pooled estimates will be inconsistent even if the relationship for each group cointegrates. The pooled estimator imposes a common slope parameter, therefore introduces an $I(1)$ component in the disturbances, thus the pooled model will not constitute a cointegrating regression. This casts doubt on the results of those studies that have pooled non-stationary series [6,12,13,16].

In practice the mean group estimator is rarely used in macroeconometrics, and there are no examples of its use in the analysis of health care spending. However, it is relatively straightforward to estimate and can be employed in a dynamic framework using a random coefficients model, such as that suggested by Pesaran and Smith [27] and Pesaran *et al.* [28]. Assuming, for simplicity, a first order heterogeneous dynamic model for health spending with only one explanatory variable, x,

$$h_{it} = \alpha_{0i} + \alpha_{1i} h_{it-1} + \beta_{0i} x_{it} + \beta_{1i} x_{it-1} + \varepsilon_{it}$$
$$i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T. \tag{3}$$

All variables are measured in log form, and all coefficients vary across countries via a random coefficients model:

$$\alpha_{1i} = \alpha_1 + \eta_{1i}, \qquad \beta_{0i} = \beta_0 + \eta_{2i}, \qquad \text{etc.}, \tag{4}$$

where η_{1i}, η_{2i}, etc., are assumed to have zero means and constant covariances. The long-run elasticity of h_i with respect to x_i is:

$$\theta_i = \frac{\beta_{0i} + \beta_{1i}}{1 - \alpha_{1i}}. \tag{5}$$

Allowing these long-run elasticities to vary randomly gives, $\theta_i = \theta + \xi_i$, where ξ_i is a stochastic term with a zero mean. The appropriate measure of the *average* long-run effect of x_i on h_i is given by the average of these long-run elasticities:

$$\bar{\theta} = \frac{1}{N} \sum_{i=1}^{N} \theta_i. \tag{6}$$

This estimator has the limit $E(\theta_i) = E(\beta_{0i} + \beta_{1i}/1 - \alpha_{1i})$, which is the same as the probability limit of the cross-section estimator.

In many data fields, including that commonly used to analyse health spending, the variables in question are $I(1)$. Pesaran and Smith [27] show that assuming a single cointegrating relationship between h_{it} and x_{it} for each country, then a static regression will provide superconsistent estimates of the long-run elasticities, which could then be averaged across countries. Assuming the x's are strictly exogenous, the cross-section estimator is also consistent in the presence of $I(1)$ variables, and the appropriate standard errors in this case are White's adjusted [29].

One problem with deriving long-run elasticity estimates from an ADL model such as (3), is that it is not straightforward to calculate a standard error for the elasticity estimate since it involves the ratio of two estimates. In order to overcome this problem, Wickens and Breusch [30] suggest a reparameterization of the ADL model, based on the Bewlay transformation [31], to provide direct estimates of the θ_i's and the associated standard errors. The reparameterization involves subtracting $\alpha_{1i} h_{it}$ from both sides, rearranging the x's and renormalizing to give:

$$h_{it} = a_{0i} + a_{1i} \Delta h_{it-1} + b_{0i} x_{it} + b_{1i} \Delta x_{it-1} + e_{it}. \tag{7}$$

The transformed variables are not independent of e_{it}, so estimation is via instrumental variables (IV), and the appropriate instruments are the explanatory variables from the ADL model (3). It then follows that $\hat{b}_{0i} = \hat{\theta}_i$, i.e. that the estimated coefficient on x_{it} is a direct estimate of the long-run elasticity of h_{it} with respect to x_{it}, and the standard error is calculated in the normal manner.

The alternative formulation of the dynamic model, (7), also has particularly useful properties when the variables in question are non-stationary and cointegrated. Estimates of the long-run effects will be consistent even if the underlying dynamic structure is misspecified by omitting difference

Copyright © 1999 John Wiley & Sons, Ltd.

terms. This is equivalent to Engle and Grangers finding that a cointegrating relationship can be obtained from a model, which is misspecified due to the omission of short-run dynamics [22].

The reparameterization suggested by Wickens and Breusch separates long run and short-run dynamics and has advantages over the ECM. The ECM restricts the long-run elasticity, in (3), to unity, i.e. $\alpha_{1i} + \beta_{0i} + \beta_{1i} = 1$, and requires the short-run dynamics to be consistent with the long-run solution. Thus, we can rewrite (3):

$$\Delta h_{it} = \beta_{0i}\, \Delta x_{it} - (1 - \alpha_{1i})(h_{it-1} - x_{it-1}) + \varepsilon_{it}. \quad (8)$$

The error correction term $(h_{it-1} - x_{it-1})$ is the short-run adjustment to equilibrium. More generally we can remove the unitary long-run elasticity, and rewrite (8):

$$\Delta h_{it} = \beta_{0i}\, \Delta x_{it} - (1 - \alpha_{1i})(h_{it-1} - \theta_i x_{it-1}) + \varepsilon_{it}. \quad (9)$$

This is simply another expression for the ADL model, that does not impose any coefficient restrictions.

Equation (9) is the specification employed by Murillo *et al.* [18], and also, presumably Saez and Murillo [19], although they do not state this. However, it is not a convenient form to estimate the long-run elasticity, θ_i, or the short-run dynamics. It is better to use the alternative formulation, (7). This model can then be subject to conventional tests of misspecification, which can confirm the statistical properties of the estimates of both the long run and short-run effects.

DATA, METHOD AND RESULTS

The dynamic heterogeneous relationship between health care spending and other potential regressors, is investigated using annual data for 20 OECD countries, 1960 to 1993. All data is from the OECD Health Data CD-ROM [32], and all variables are in natural logarithms. In order to facilitate comparisons with existing studies the choice of regressors has been made on the basis of past literature and a detailed justification will not be repeated here (see Gerdtham *et al.* [9] for a good discussion).

h is health care expenditure and y is GDP, both are per capita, measured in constant US dollar prices, and adjusted via purchasing power parities. p is the percentage of publicly funded health care spending; there is contrary theory and evidence in relation to both the size and direction of its effects on total health spending. a is the percentage of the population aged over 65 years, included to account for the fact that ageing populations generally use more health resources. r is the relative price of health care, i.e. the ratio of the health price index to the GDP deflator. This is included to separate price and income effects and see if richer countries actually purchase more resources or simply pay higher factor prices.

The time trend, t, is a simple proxy for technical change in the absence of any other usable measure. Technical progress may reduce the cost of producing a given output, however, it also increases the effectiveness of medical interventions producing both demand and supply side reasons for increasing expenditure. In reality the time trend will account for any non-observed trended effects, and its coefficient is unlikely to be robust due to collinearity with other trended variables.

The models of health spending adopted in the literature are *ad hoc*, and this has implications for the econometric analysis. A weak theoretical base gives little guidance as to the possible explanatory variables and the causal mechanisms involved. Traditionally econometric results are judged against three sets of criteria. Theoretical criteria involve the comparison of results with a priori expectations. Statistical criteria involve tests for individual and joint significance of parameter estimates. Econometric criteria involve tests of the assumptions underlying estimation, and enable evaluation of the properties of the estimators. A lack of theoretical guidance necessarily increases the emphasis on statistical and econometric criteria, and ideally results should be subject to stringent sensitivity analysis in order to establish robustness in the absence of theoretical expectations. Unfortunately this has not been the case with much of the existing literature.

In preparation for structural modelling all the variables have been tested for unit roots using augmented Dickey–Fuller (ADF) tests [33], Phillips–Perron tests [34] and the 't-bar test' [35], which exploits the panel structure of the data. All results are sensitive to the inclusion of a time trend in the testing equations. When a time trend is included, the results are robust to the number of lags and to demeaning (which minimizes the effect of structural breaks and inappropriately specified trends), and provide virtually no evidence

Copyright © 1999 John Wiley & Sons, Ltd.

to reject the null hypothesis of a unit root, for any variable for any country. Without a time trend the unit root null is rejected for h and p in almost half of the countries (and for the t-bar test), although not for the demeaned series. The presence of a time trend in the underlying series was investigated using the sequential testing procedure suggested by Holden and Perman [36]. This confirmed that most series were characterized by both a unit root and a linear trend. Overall the above evidence seems to suggest that for all countries the variables are $I(1)$, and these results confirm those of previous studies [16,23]. However, in the analysis that follows particular attention will be paid to the sensitivity of the results to the time trend.

A first order ADL model with random coefficients forms the basis of analysis:

$$h_{it} = \alpha_{0i} + \alpha_{1i}h_{it-1} + \alpha_{2i}t + \beta_{0i}y_{it} + \beta_{1i}y_{it-1} + \delta_{0i}p_{it} + \delta_{1i}p_{it-1} + \gamma_{0i}a_{it} + \gamma_{1i}a_{it-1} + \lambda_{0i}r_{it} + \lambda_{1i}r_{it-1} + \varepsilon_{it} \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T. \tag{10}$$

All coefficients vary randomly across countries, i, as defined in (4).

Table 1 reports the long-run elasticity estimates from a variety of estimators. Those in the first column are obtained from the Wickens and Breusch reparameterization of (10) estimated via IV. They are the means of the country specific long-run estimates defined by (6). These estimates are superconsistent when the variables in question are $I(1)$. The income elasticity estimate is larger than usually reported in the literature. The other regressors do not have significant long-run effects.

Column (2) of Table 1 shows the long-run estimates derived from a static model. The income elasticity is much lower than that obtained from the dynamic model but is still significantly greater than unity at $t_{0.01}$. In addition p and r both have significant long-run effects. However, the standard errors may be misleading because they are based on inappropriate estimates of the variance. The properties of these estimates depend on whether this model actually represents a cointegrating relationship, and evidence is inconclusive. Using the Engle–Granger procedure [22], the null of non-stationary residuals, i.e. no cointegration, is rejected in only four cases. Using the t-bar test [35] to exploit the panel structure of the data, gives a statistic of -3.712, which rejects the unit root null at the 1% level. The multivariate Johansen procedure [37] finds at least one cointegrating vector for all but one country. The Pesaran *et al.* test, which is valid when the order of integration of the underlying regressors is unknown, only rejects the null of no long-run relationship in five cases [38].

Estimates for the pooled model with fixed effects are given in the second section of Table 1. The estimates in (3) are derived from a dynamic

Table 1. Long-run elasticity estimates from alternative estimators

	Mean group		Pooled		Cross section
	(1)[a]	(2)[a,b]	(3)	(4)	(5)[c]
y	2.041*	1.254*	1.875	1.515*	1.514*
	(0.428)	(0.079)		(0.056)	(0.161)
p	0.704	0.759*	0.523	0.417*	−0.128
	(0.405)	(0.089)		(0.052)	(0.214)
a	−0.895	−0.397	−0.830	−0.019	0.060
	(0.698)	(0.224)		(0.080)	(0.220)
r	0.215	0.332*	0.466	0.329*	−0.427
	(0.249)	(0.076)		(0.045)	(0.399)

All regressions for $i = 20$, $t = 34$ (1960–1993). (1) and (3) ADL models, (2) and (4) static models.
* Statistically significant at $t_{0.05}$.
[a] Standard errors for the mean group estimator are computed assuming that regression coefficients are independently distributed across countries.
[b] Standard errors for the cointegrating regressions are based on inappropriate estimates of the variance.
[c] White's adjusted standard errors.

Copyright © 1999 John Wiley & Sons, Ltd.

Table 2. Distribution of estimates underlying the mean group estimates

	y	*p*	*a*	*r*	*t*
Mean	2.041*	0.704	−0.895	0.215	−0.024
(standard error)	(0.428)	(0.405)	(0.698)	(0.215)	(0.014)
Median	1.380	0.516	−0.583	0.277	−0.014
Standard deviation	2.180	2.134	2.877	0.954	0.075
Minimum	−2.210	−3.515	−6.025	−1.395	−0.194
Maximum	7.190	6.109	2.826	2.152	0.133

* Statistically significant at $t_{0.05}$.
Estimates correspond to Table 1, column (1).

model, and those in (4) are from the static model. These estimates are not consistent but are provided for comparison, since they are commonly employed in the health expenditure literature. These results are similar to those derived from the mean group estimators. However, the static model will not constitute a cointegrating relationship even if individual country relationships cointegrate, due to the imposition of a common cointegrating vector across countries.

The last column, (5), gives the results from a cross-section regression using averages for 1960 to 1993. This model provides consistent estimates of the long-run effects. However, compared with the estimates in columns (1) and (2), apart from the income elasticity, the cross section does not provide good point estimates of the mean of the long-run coefficients. This may be due to interaction effects between the groups. The long-run income elasticity is significantly greater than one, but for all the other coefficients the signs are reversed, although none of these estimates are statistically significant.

Overall the results are fairly similar across the estimators. All of the estimates of the long-run income elasticity of health spending are significantly greater than one, and the signs of the other long-run effects are quite robust to model specification, except in the cross-section case. Of the consistent estimators, the mean group estimator is preferred to the cross section because heterogeneity is explicit. All of the random coefficients models were tested against the null hypothesis of no slope variation and this was rejected in every case. Despite the contradictory evidence on cointegration, there is little to choose between the results from the static and dynamic models. This robustness across specification may lead us to have more confidence in the results.

It appears that the long-run income elasticity of health spending is greater than one. The positive coefficient on *p*, suggests that a 10% increase in the share of public spending, increases total health spending by about 7%; a result similar to that found in Leu's early cross-section study [10]. The relative price effect is positive suggesting that total health demand is not constrained by price increases, as would be the case in a fixed budget regime. However, this coefficient is not statistically significant in all of the models. The effect of the ageing population is not significant, and the negative sign is contrary to expectations.

The individual country regressions were tested for heteroscedasticity using White's test [29], autocorrelation using the Breusch–Godfrey test [39], and normality of the residuals using the Jarque–Bera test [40]. Stability of the estimates was also tested using the Chow forecast test with the sample split in 1975 [41]. Overall the underlying regressions appear to be well specified, and this confirms the statistical properties of the long-run estimates. The few specification problems that do exist are associated with four countries for which the first order ADL model is rejected as a restriction on a more general model. For Canada and France the problems are solved with a second order ADL model, and for Norway by including a second lagged dependent variable. As predicted by the theory, these changes make little difference to the estimated long-run elasticities for these countries. It proved impossible to find a model for Belgium that satisfied all of the diagnostic tests.

Copyright © 1999 John Wiley & Sons, Ltd.

SENSITIVITY ANALYSIS

The heterogeneity of the relationships underlying the mean group estimates is shown in Table 2, which reports the distribution of estimates across countries. All distributions have large standard deviations, and span the negative/positive range. The distribution of the long-run income elasticity estimate is highly skewed, and this is caused by five particularly high estimates for Belgium, Germany, Ireland, Italy and the Netherlands. In contrast to the findings of Saez and Murillo [19] there does not appear to be any relationship between the estimated income elasticity and the level of income in a country (correlation coefficient of -0.27). Despite the heterogeneity, the average long-run estimates are robust to the sample of countries used in the averaging process. This is illustrated in Figure 1(a), which shows the effect of excluding each country in turn on the average estimate of the elasticities.

Parameter stability through time is investigated by estimating the cross section recursively. Plots of these recursive estimates, shown in Figure 2, reveal systematic variation over the time period. The income elasticity estimate seems to fall throughout the 1970s and has been rising since the late 1980s. Interestingly, there does not appear to be a significant effect from the severe recession of the mid-1970s. The most striking result is that for the long-run relative price effect (r) which has been falling throughout the period from 0.245 to -0.427.

The results are extremely sensitive to the exclusion of a time trend, as shown in Table 3. The first column (1a) shows the dynamic model estimated without a time trend. None of the estimates are very precise and the reason for this is illustrated by Figure 1(b), which shows the sensitivity of this model to the cross section of countries used in the averaging process. All the estimated average long-run effects are extremely sensitive to the exclusion of Italy. The second column of

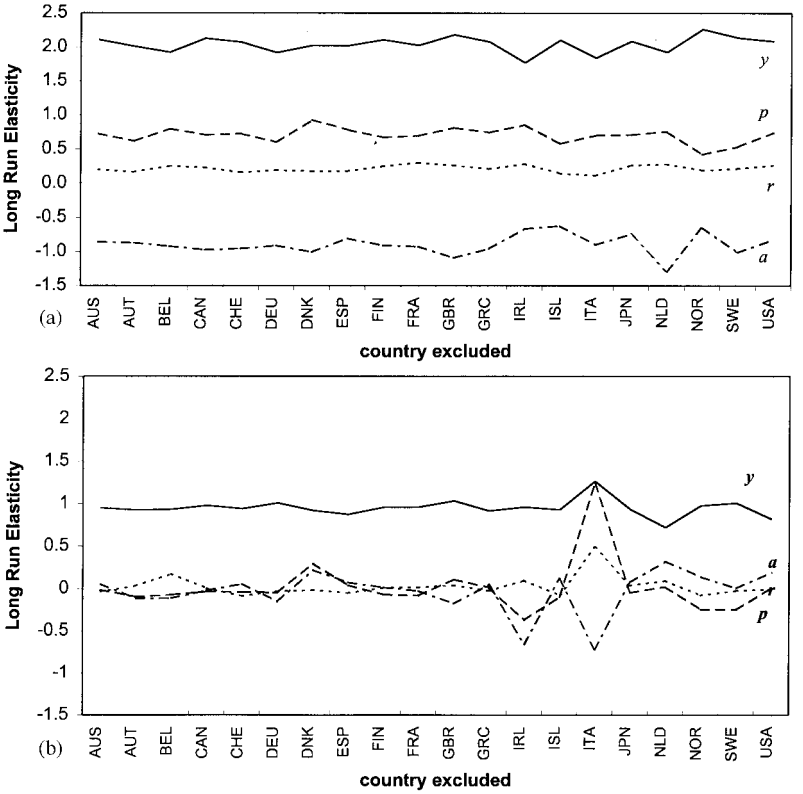

Figure 1. Dynamic model: (a) cross-section sensitivity with T; (b) cross-section sensitivity with no T

Copyright © 1999 John Wiley & Sons, Ltd.

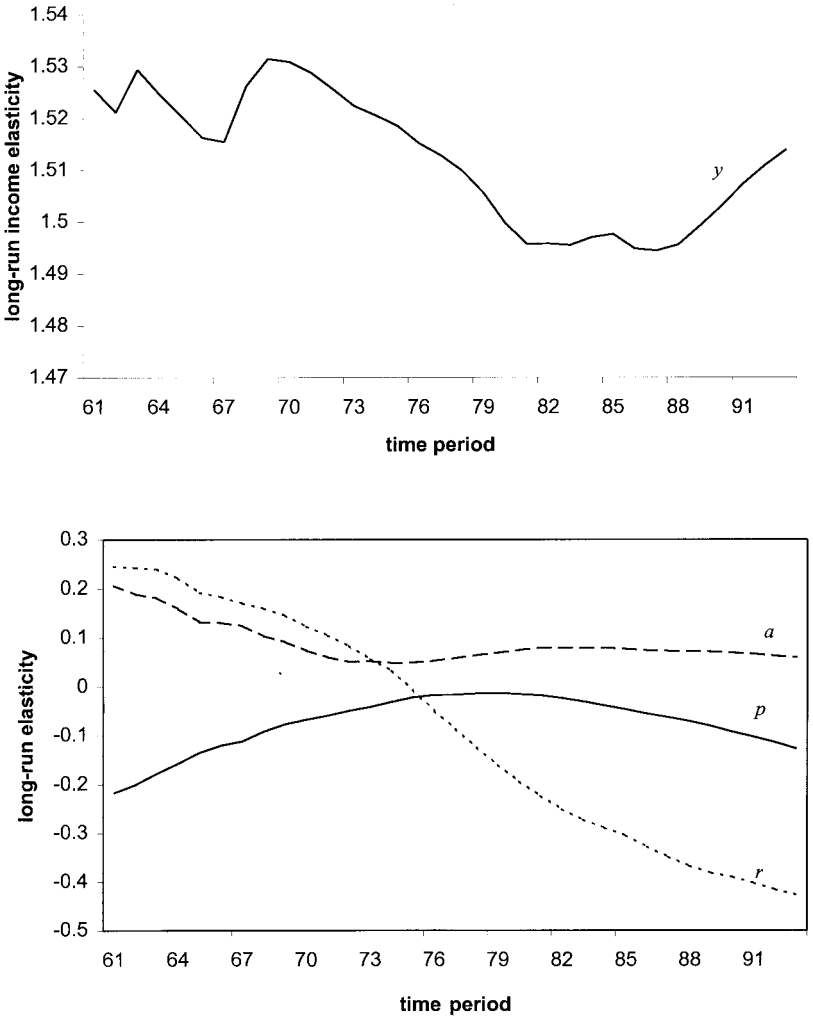

Figure 2. Recursive estimates of long-run coefficients

Table 3, (1b), shows the results for the same model when Italy is excluded, and these results are much more in line with those reported in Table 1, column (1).

The question of whether or not to include a time trend in the dynamic model is obviously crucial. The trend term is only statistically significant ($t_{0.05}$) in five of the 20 individual country regressions. However, as Table 2 shows, the mean estimate is significant at $t_{0.10}$, and it suggests that once differences in income, public spending, age and relative prices are taken into account, the average tendency is for health spending to fall by 2.4% per year. The insignificance of t in the underlying regressions, combined with disruption to the other coefficient estimates when it is excluded, is suggestive of collinearity between the trend term and the other variables measured in levels, since these also contain deterministic, as well as stochastic, trends. Pairwise correlation coefficients confirm this.

The problem with excluding t is that the results become very sensitive to sample composition and model specification. In contrast when t is included, the results are more robust, as shown in the second section of Table 3. Comparing these results to the Table 1, column (1), it seems that only the omission of p has any major effect. Wald tests confirm that both y and p should be included in nearly all of the underlying regressions. The

Copyright © 1999 John Wiley & Sons, Ltd.

problem with simplifying the equations by excluding a and r is that some of the models become very unstable, and as a result the average estimates become very sensitive to the cross section of countries used. It is therefore safer to include all of the variables, including t, due to their theoretical importance, the cross-sectional stability of the estimates and the statistical properties of the models.

Finally, we should consider the short-run dynamics implied by the first order ADL, and also consider dynamic simplifications. Short-run dynamics are given by the coefficients on the differenced variables in the reformulated version of (10). The coefficients on Δy, Δp, Δa and Δr, respectively, are 0.017, -0.459, 1.581 and 1.036. Only the latter is statistically significant at $t_{0.10}$, and unlike the long-run estimates, the short run effects are very sensitive to the cross section of countries used in the averaging process. Nevertheless they do make sense intuitively, suggesting that the short-run effect of changes in income, public spending and age composition are small, or non-existent, compared with the long-run effect. However, the short-run effect of relative prices is more important than the long-run effect and suggests that in the short-run price increases are met by a one-for-one increase in health expenditure.

Simplifications to the dynamic structure are tested, via Wald tests, as restrictions on Equation (10). The restrictions and results are summarized in Table 4. The static model, that has characterized much of the literature, is rejected for all countries. The strictest ECM(1) that forces all coefficients to sum to 1, is rejected in only five of the 20 cases. The less restrictive ECM(2) that restricts only the income elasticity (to 1), appears satisfactory for all countries. Although, there is now some conflict with the unrestricted model which predicts an income elasticity greater than one. Gerdtham's [6] model, ECM(3), is rejected in nearly all cases.

Estimates from the chosen ECM give an average adjustment coefficient of -0.221, which is statistically significant at $t_{0.01}$. This suggests that countries take an average of about 5 years to adjust to long-run equilibrium, and this adjustment is slower than predicted by Saez and Murillo [19] for the EC countries, but similar to the findings of Gerdtham [6] for the OECD. The error correction term is sensitive to the cross section of countries used, but it is has the correct negative sign for all countries except Italy, and is statistically significant in 12 cases. This adds support to the idea of an equilibrium relationship between y and h, despite contradictory results from the cointegration tests.

CONCLUSION

This paper has considered the difficulties of estimating elasticities of health care spending in the light of four factors. Firstly, the relationships involved are heterogeneous across countries. Secondly, the relationships are characterized by dynamic effects. Thirdly, the variables involved are non-stationary. Fourthly, theory gives very little guidance against which to judge the estimates.

Table 3. Sensitivity analysis of mean group estimates

	Time trend excluded		Exclusion of explanatory variables		
	(1a)	(1b)	(2)	(3)	(4)
y	0.948	1.263*	3.180*	2.103*	2.103*
	(1.877)	(0.348)	(0.882)	(0.870)	(0.383)
p	−0.226	1.233		0.735	0.749*
	(6.715)	(0.911)		(0.488)	(0.367)
a	0.660	−0.720	−1.288		−1.120
	(4.217)	(1.942)	(1.484)		(0.594)
r	−0.514	0.487	−0.117	0.315	
	(2.542)	(0.389)	(0.571)	(0.369)	

See notes to Table 1.
(1a) ADL model, (1b) as (1a) but excluding Italy. (2)–(4) include a time trend.

Copyright © 1999 John Wiley & Sons, Ltd.

Table 4. Restriction tests on ADL mean group estimator

Model	Restriction	No of rejections (from 20)
Static	$\alpha_1 = \alpha_2 = \beta_1 = \delta_1 = \gamma_1 = \lambda_1 = 0$	20
Partial adjustment	$\alpha_2 = \beta_1 = \delta_1 = \gamma_1 = \lambda_1 = 0$	10
Growth rate	$\alpha_2 = 0,\ \alpha_1 = 1,\ \beta_0 = -\beta_1,\ \delta_0 = -\delta_1,\ \gamma_0 = -\gamma_1,\ \lambda_0 = -\lambda_1$	17
ECM(1)	$\alpha_2 = 0,\ \alpha_1 + \beta_0 + \beta_1 + \delta_0 + \delta_1 + \gamma_0 + \gamma_1 + \lambda_0 + \lambda_1 = 1$	5
ECM(2)	$\alpha_2 = 0,\ \alpha_1 + \beta_0 + \beta_1 = 1$	0
ECM(3)—Gerdtham [6]	$\alpha_2 = 0,\ \alpha_1 + \beta_0 + \beta_1 = 1,\ \delta_0 = -\delta_1,\ \gamma_0 = -\gamma_1,\ \lambda_0 = \lambda_1 = 0$	15

These represent tests of restrictions on the ADL model (10) estimated for each of 20 OECD countries: $h_{it} = \alpha_{0i} + \alpha_{1i} h_{it-1} + \alpha_{2i} t + \beta_{0i} y_{it} + \beta_{1i} y_{it-1} + \delta_{0i} p_{it} + \delta_{1i} p_{it-1} + \gamma_{0i} a_{it} + \gamma_{1i} a_{it-1} + \lambda_{0i} r_{it} + \lambda_{1i} r_{it-1} + \varepsilon_{it}$.

Using techniques designed to cope with the first three of these factors, the results suggest that long-run elasticity estimates are sensitive to model specification, and in particular to the exclusion of a time trend. With a trend term, the long-run income elasticity of health spending is greater than one, but without it is approximately unity. However, the omission of the trend results in unstable parameter estimates and considerable sensitivity to the cross-section composition of the sample. It is unclear theoretically, whether a trend should be included in the estimating equations, and this theoretical weakness (the fourth factor) creates by far the biggest problems for estimation and evaluation.

Attempts have been made here to compensate for the weak theoretical framework by increasing the emphasis on statistical and econometric criteria to judge the validity of the results. However, the sensitivity of the results to model specification and sample composition only serves to increase the need for a stronger theoretical base. Despite these theoretical problems, some results warrant further mention. Once again income has emerged as the primary correlate of health spending, and the tests of short-run dynamics seem to suggest that the long-run income elasticity is approximately unity. Public spending and age composition do not seem to have any short-run effects on health spending, but the proportion of public spending does have a significant positive long-run effect in the cointegrating model. Short-run changes in relative prices are important and result in a one-for-one increase in health spending. The relative price and public spending estimates both suggest that there is a role for policy in controlling health spending.

The primary aim of future work must be in strengthening the theoretical base from which to analyse health spending. This could build on the work of Dunne and Smith [42], Dunne *et al*. [43] and Selvanathan and Selvanathan [44], who analyse health spending within a macroeconomic framework of expenditure or consumption behaviour.

Empirically the analysis could be improved by extending the data set to lower income countries, and attempts have already been made in this area [45,46]. In addition, while this paper has assumed that the regression coefficients are independently distributed across countries, it would also be useful to consider estimation techniques that combine parameter heterogeneity with cross country correlations and interaction effects. Finally, Baltagi and Griffin [47] have recently questioned the forecasting performance of the mean group estimator, particular in relation to long range forecasts, and suggest that the efficiency gains from pooling could more than offset the biases due to heterogeneity. This issue should be investigated empirically within the context of health care expenditure.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the Seventh European Workshop on Econometrics and Health Economics (Helsinki, Finland, September 1998). I would like to thank participants at the workshop for their useful suggestions and criticisms, and particularly Iris Frohwitter for her discussant comments.

REFERENCES

1. Newhouse, J.P. Medical care expenditure a cross-national survey. *Journal of Human Resources* 1977; **12**: 115–125.

Copyright © 1999 John Wiley & Sons, Ltd.

2. McGuire, A., Parkin, D., Hughes, D. and Gerard, K. Econometric analyses of national health expenditures: can positive economics help to answer normative questions? *Health Economics* 1993; **2**: 113–126.
3. Quah, D. *International patterns of growth I: Persistence in cross-country disparities*. Mimeo, Economics Department, MIT, Cambridge, MA, 1990.
4. Organisation for Economic Co-operation and Development. *Health care systems in transition: the search for efficiency*. Social Policy Studies No. 7. Paris: OECD, 1990.
5. Parkin, D., McGuire, A. and Yule, B. Aggregate health care expenditures and national incomes: is health care a luxury good? *Journal of Health Economics* 1987; **6**: 109–127.
6. Gerdtham, U. Pooling international health care expenditure data. *Health Economics* 1992; **1**: 217–231.
7. Abel-Smith, B. *Cost containment in healthcare*. London: Bedford Square Press, 1984.
8. Organisation for Economic Co-operation and Development. *Financing and delivering healthcare*. Social Policy Studies No. 4. Paris: OECD, 1987.
9. Gerdtham, U., Sogaard, J., Andersson, F. and Jonsson, B. An econometric analysis of health care expenditure: a cross-section study of the OECD countries. *Journal of Health Economics* 1992; **11**: 63–84.
10. Leu, R.E. The public–private mix and international health care costs. In: Culyer, A.J. and Jonsson, B. (eds.), *Public and private health services: complementarities and conflicts*. Oxford: Blackwell, 1986.
11. Hsiao, C. *Analysis of panel data*. Cambridge: Cambridge University Press, 1986.
12. Hitiris, T. and Posnett, J. The determinants and effects of health expenditure in developed countries. *Journal of Health Economics* 1992; **11**: 173–181.
13. Hitiris, T. Health care expenditure and integration in the countries of the European Union. *Applied Economics* 1997; **29**: 1–6.
14. Roberts, J. Spurious regression problems in the determinants of health care expenditures: a comment on Hitiris. *Applied Economics Letters* 2000; forthcoming.
15. O'Connell, J.M. The relationship between health expenditures and the age structure of the population in OECD countries. *Health Economics* 1996; **5**: 573–578.
16. Blomqvist, A.G. and Carter, R.A.L. Is health care really a luxury? *Journal of Health Economics* 1997; **16**: 207–229.
17. Pfaff, M. Differences in health care spending across countries: statistical evidence. *Journal of Health Politics, Policy and Law* 1990; **15**: 1–67.
18. Murillo, C., Piatecki, C. and Saez, M. Health care expenditure and income in Europe. *Health Economics* 1993; **2**: 127–138
19. Saez, M. and Murillo, C. Shared 'features' in prices: income and price elasticities for health care expenditures. *Health Economics* 1994; **3**: 267–281.
20. Getzen, T.E. and Poullier, J. International health spending forecasts: concepts and evaluation. *Social Science and Medicine* 1992; **34**: 1057–1068.
21. Kendix, M. and Getzen, T.E. US Health services employment: a time series analysis. *Health Economics* 1994; **3**: 169–181.
22. Engle, R.F. and Granger, C.W.J. Cointegration and error correction: representation, estimation and testing. *Econometrica* 1987; **55**: 251–276.
23. Hansen, P. and King, A. The determinants of health care expenditure: a cointegration approach. *Journal of Health Economics* 1996; **15**: 127–137.
24. McCoskey, S.K. and Selden, T.M. Health care expenditures and GDP: panel data unit root tests results. *Journal of Health Economics* 1998; **17**: 369–376.
25. Hansen, P. and King, A. Health care expenditures and GDP: panel data unit root tests results—comment. *Journal of Health Economics* 1998; **17**: 377–381.
26. Zellner, A. On the aggregation problem: a new approach to a troublesome problem in economics. In: Fox, K.A. *et al*. (eds.), *Economic models, estimation and risk programming: essays in honor of Garhard Tintner*. Berlin: Springer-Verlag, 1969.
27. Pesaran, M.H. and Smith, R. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 1995; **68**: 79–113.
28. Pesaran, M.H., Smith, R. and Im, K.S. Dynamic linear models for heterogeneous panels. In: Matyas, L. and Sevestre, P. (eds.), *The econometrics of panel data: handbook of theory with applications*, 2nd edn. Amsterdam: Kluwer Academic, 1995.
29. White, H. A heteroscedasticity consistent covariance matrix and a direct test for heteroscedasticity. *Econometrica* 1980; **48**: 817–838.
30. Wickens, M.R. and Breusch, T.S. Dynamic specification, the long-run and the estimation of transformed regression models. *Economic Journal* 1988; **98**(Supplement): 189–205.
31. Bewlay, R.A. The direct estimation of the equilibrium response in a linear model. *Economics Letters* 1979; **3**: 357–361.
32. Organisation for Economic Co-operation and Development. *OECD Health Data 1996.* Paris: OECD, 1996.
33. Dickey, D.A. and Fuller, W.A. Distribution of the estimators for autoregressive time-series with a unit root. *Journal of the American Statistical Association* 1979; **74**: 427–431.
34. Phillips, P.C.B. and Perron, P. Testing for a unit root in time series regression. *Biometrika* 1988; **75**: 335–346.

Copyright © 1999 John Wiley & Sons, Ltd.

35. Im, K.S., Pesaran, M.H. and Shin, Y. *Testing for unit roots in heterogeneous panels*. DAE Working Papers 9526, University of Cambridge, 1995.
36. Holden, D. and Perman, P. Unit roots and cointegration for the economist. In: Rao, B.B. (ed.), *Cointegration for the applied economist*. Basingstoke: St Martins Press, 1994.
37. Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 1991; **59**: 1551–1580.
38. Pesaran, M.H., Shin, Y. and Smith, R. *Testing for the existence of a long-run relationship*. DAE Working Papers 9622, University of Cambridge, 1996.
39. Godfrey, L.G. *Specification tests in econometrics*. Cambridge: Cambridge University Press, 1988.
40. Bera, A. and Jarque, C. Efficient tests for normality, heteroscedasticity and serial independence of regression residuals. *Economics Letters* 1980; **6**: 255–259
41. Chow, G. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 1960; **28**: 591–605.
42. Dunne, J.P., Parshardes, P. and Smith, R. Needs costs and bureaucracy: the allocation of public consumption in the UK. *Economic Journal* 1984; **94**: 1–15.
43. Dunne, J.P. and Smith, R. The allocation of efficiency of government expenditure: some comparative tests. *European Economic Review* 1983; **20**: 381–394.
44. Selvanathan, S. and Selvanathan, E.A. A cross country analysis of consumption patterns. *Applied Economics* 1993; **25**: 1245–1259.
45. Gbesemete, K.P. and Gerdtham, U. Determinants of health care expenditure in Africa: a cross section study. *World Development* 1992; **20**: 303–308.
46. Cullis, J.G. Medical care expenditure: a cross national survey: an extension to developing nations. Mimeo, University of Bath, 1977.
47. Baltagi, B.H. and Griffin, J.M. Pooled estimators vs. their heterogenous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 1997; **77**: 303–327

Copyright © 1999 John Wiley & Sons, Ltd.